SSOAR

Open Access Repository
www.ssoar.info

Efficiency of hospitals in Germany: a DEA-bootstrap approach

Staat, Matthias René

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Staat, M. R. (2006). Efficiency of hospitals in Germany: a DEA-bootstrap approach. *Applied Economics*, *38*(19), 2255-2263. https://doi.org/10.1080/00036840500427502

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-239226

Efficiency of hospitals in Germany: a DEA-bootstrap approach

Journal:	*Applied Economics*
Manuscript ID:	APE-05-0018.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	01-Jun-2005
JEL Code:	C44 - Statistical Decision Theory|Operations Research < C4 - Econometric and Statistical Methods: Special Topics < C - Mathematical and Quantitative Methods, I12 - Health Production: Nutrition, Mortality, Morbidity, Substance Abuse and Addiction, Disability, and Economic Behavior < I1 - Health < I - Health, Education, and Welfare
Keywords:	Data Envelopment Analysis, Hospitals Efficiency , Bootstrapping

powered by ScholarOne
Manuscript Central™

I INTRODUCTION

During the 1990s health related expenditures in Germany amounted to more than 10% of GDP; [1] the largest single item among all outlays was in-patient care, which added up to around 3.5% of the GDP. In 1995, this equalled € 45bn (OECD, 2001).[2] Therefore, cost containment in the hospital sector is a key issue in stabilising health related expenditures at a sustainable level.

Several studies comparing health care systems internationally concluded that the German system is not efficiently managed.[3] Especially the overcapacity of hospitals beds in Germany is considered to be a source of inefficiency. There are numerous reasons behind the excess –the OECD (1997) estimates 14% but much higher estimates exist - of beds in Germany implied by these figures.

A number of studies on the efficiency of hospitals in Germany attempted to quantify the savings potential. Neither of these studies gives a result representative for all German hospitals or for a specific segment of hospitals. For instance Henke, Wettke and Paffrath (1995) assess the cost efficiency of German hospitals by comparing the average case cost of hospitals in different cities. They give examples of two cities where cost for treating one particular ICD exceed the national average by 20% and 53%, and conclude that there are „dramatic differences in efficiency" between German hospitals.

Swart et al. (1996) derive a ranking for 50 hospitals which treated patients insured with the Magdeburg (Saxony-Anhalt) regional subdivision of the Allgemeine Ortskrankenkasse (AOK), the leading German sickness fund. They rank hospitals by their length of stay (LOS) for the most common ICDs. The ICDs receive a number of points equal to #observations + 1 – rank; these points are then summed and divided by the maximum number of points which results in a score between 0 and 1. The scores for a subsample of nine homogeneous hospitals with a focus on internal medicine range between .77 and .27. According to the authors, these differences cannot be explained by

For Peer Review

factors such as differing age of patients; consequently, they consider this variation as an indication of a substantial savings potential.

Recently, a few studies on the relative efficiency of German hospitals using Data Envelopment Analysis (DEA) have been published (for surveys on health care studies using DEA, see Hollingsworth, Dawson and Maniadakis, 1999 and Chilingerian, 2004). Steinmann et al. (2004) report results for the federal state of Saxony for the years 2000 to 2002. Their sample comprises 105 hospitals with at least 20 beds which do not provide tertiary care. They assume constant returns to scale (CRS) and use academic, administrative and nursing staff as well as other expenses, patient days and beds as inputs. The number of cases treated in the internal medicine, paediatric, gynaecological, surgical or intensive care department function as (separate) outputs. The average efficiency for these hospitals ranges between 79 % (2002) and 83 % (2000).

Kuntz and Scholtes (2004) report results on 92 hospitals located in Rhineland-Palatinate based on data from 2001. They assume a CRS technology and employ two different model specifications.[4] Their input parameters are the number of beds and total cost. One model is based on the total number of cases in 20 different ICD-clusters (model 1) as outputs; the other contains patients who did not stay over night and were treated in one of 20 different departments (model 2). Model 1 results in an average efficiency score of 95 % with 72.83 % of the observations considered to be fully efficient. Less than 8 % of the hospitals have an efficiency score below 90 %.

Both DEA studies on Germany hospitals are based on a regional selection comprising very different hospitals in terms of departments and specialization. Therefore, one may expect their results to be downward biased because samples with similar hospitals would generate results that would have a higher average efficiency. On the other hand, DEA results are biased upward (see section III for details) and it is not possible to tell which of the two effects dominates.

In the sequel, a DEA study on German hospitals will be carried out employing a precursor data set to the one collected for the hospitals benchmark, which was introduced into Germany law in 1998.[5] Unlike other DEA studies, the results derived are representative for hospitals in the old federal states in Germany.

This paper is organised as follows: in the next section the data are described in detail. This is followed by an introduction to the bootstrapping method employed for the empirical analysis.[6] Next, the results of the analysis are presented and their implications are discussed. A brief summary of the findings concludes.

II DATA

The data used in the present study are a precursor data set to the one used for the mandatory hospital benchmark in Germany. This benchmark is carried out at the Wissentschaftliches Institut der Ortskrankenkassen (WIdO), the research institute of the AOK. The data are from 1994.[7] These are the latest data available that allow for a consistent benchmarking of hospital performance. Hospitals were remunerated with a base per diem that applied for all departments of the hospital. This negotiated per diem is available in the data contained in Arnold and Paffrath (1996).

The information[8] for some 1700 hospitals includes type of ownership (public, private or owned by a non-profit organisation), size (number of beds) and structure of the hospital (number and type of departments). For the following five departments, the average LOS and the case mix cluster (see the explanation below) are known: internal medicine, surgery, gynaecology, orthopaedics and ENT.

It is obvious that the entire services a hospital provides are reflected by its per diem. Hence, only hospitals with a similar range of services can be compared. Hospitals are grouped into four categories by their function in German health care: hospitals of only local importance providing basic care without any large scale technical facilities (type I), basic care hospitals with some facilities that are of regional importance (type II),

For Peer Review

hospitals with several departments which are of central importance for the region (type III) and tertiary care hospitals (type IV). In order to ensure general comparability of hospitals, the analysis is restricted to type I, local, and type II, regional hospitals.

Also, the information on structural classes (Strukturgruppen – hospitals with a similar departmental structure) is used. The concept of structural classes was developed for the purpose of identifying "comparable" hospitals; the information on structural classes is contained in the data supplied with Arnold and Paffrath (1995).[9] There are 77 different classes and it is suggested that "only hospitals belonging to the same structural class should be compared …" (Arnold and Paffrat, 1995, p. 273, translation by the author).

Only groups 11 to 13 of the 77 structural groups are included in the analysis. All these hospitals provide basic care. They have two main departments: one for medicine and one for surgery. The only difference between the three structural classes is the fraction of beds which are made available for patients treated by external specialists.[10] This leaves 160 hospitals, 108 of type I and 52 of type II.[11] Summary statistics on the hospitals in our data are given in Table 1.

Table 1: Summary statistics

One input indicator used to compare these hospitals is their per diem rate. This differs from the WIdO-Benchmark, which is based on departmental case cost. A specification that treats the per diem rate and LOS separately is preferred as this makes it possible to analyse the efficiency of the hospitals in more detail.[12]

The second input is the number of beds. Since there may be economies of scale and/or differences between the learning rates of institutions of different size it is necessary to ensure that the individual hospital and the benchmark technology are comparable in this respect. As the number of beds cannot be changed in the very short run it is treated as a non-discretionary variable (see Banker and Morey, 1986, and Staat 1999).

For Peer Review

The number of cases per year and the reciprocal of LOS, R-LOS, in the two fields of specialisation, model output.[13] R-LOS is used to ensure that hospitals with a low per diem but high LOS do not appear efficient or conversely hospitals with a higher per diem but low LOS do not appear inefficient.

Also, departments with an adverse case mix should only be compared to other departments with a like case mix. Departments with an adverse case mix cannot be expected to have the same output-to-cost ratio as departments with an ordinary case mix. Departments treating ordinary case mix can be expected to have no higher resource use than other departments, however. Therefore, an indicator for adverse case mix is used as an additional output descriptor.

Gerste (1996) divides surgery departments into six case mix clusters (for a description of the case mix clusters, see Gerste, 1996, Table 2). Only one of the clusters differs substantially from the others w. r. t. either LOS, intensive care days per 100 cases, minutes of care per day, the fraction of patients older than 75 years or the per diem rate. The cluster with the most heterogeneous case mix has about twice as many intensive care days (59.1) as the remaining 5 clusters (between 22.6 and 33.1 days). The fact that this is the only cluster with cases that may require significantly more resources than the patients in any of the other clusters is reflected in the comments made by Gerste (1996, p. 123; translation by the author): "Although hospitals treat different case mixes there is no difference in the services they provide. ... Only the heterogeneous class with its high values for intensive care differs from the others. “

Of the four case mix clusters into which the internal medicine departments are divided again only the most heterogeneous group with an average LOS of 16.1 days differs from the other three groups for which LOS ranges from 12.9 to 13.8 days.[14] An output dummy for adverse case mix is set to one for the internal medicine or surgery departments with a heterogeneous case mix.

For Peer Review

Because the hospitals in the data belong to the segment of the hospital system that provides the most basic form of clinical care and involved cases are likely to be referred to more specialised hospitals only very few hospitals face an adverse case mix (see Table 1). A significant difference for the proportion of departments with an adverse case mix indicates that surgery departments in type II-hospitals treat a more heterogeneous case mix than departments in type I-hospitals as was to be expected.

Table 2: Characteristics of case mix clusters

Hospitals may admit patients who are treated by external specialists of various disciplines. Hospital management will require the more resources the more disciplines; in order to take this potentially efficiency relevant factor into account, an indicator for the number of all fields of specialisation including those represented by external specialists is used.

The two input (discretionary: per diem; nondiscretionary: beds) and six output indicators (cases, R-LOS and case mix for the medicine and the surgery department as well as fields of specialisation) described above are used to assess the efficiency of the hospitals.

Table 1 contains summary statistics of the sample. On average, hospitals of type II have a slightly lager number of beds than hospitals of type I but the smallest hospital in the sample is of type II. The latter hospitals on average have slightly more fields of specialisation than type I hospitals but the hospital with the most fields of specialisation is of type I (see Table 1).

Except for LOS in the surgery department, which takes a maximum of three weeks in the local and of only two weeks in the regional hospitals, all indicators have very much the same range for the two types of hospitals. Treatment duration is longer for the inter-

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

For Peer Review

nal medicine departments and shorter for the surgery departments of the type I hospitals when compared to type II hospitals. These differences are not significant, however.

The parameters used in the analysis ensure that the hospitals to be compared will be very similar w. r. t. structure, size, case mix and tasks performed. The effects of any remaining differences on the results will be further reduced by the way in which DEA constructs the reference technologies. Since the observations will be compared only to reference technologies with an almost identical input-output-structure (see formula (4) below) the results should only reflect differences in the efficiency of service provision.

III DEA & BOOTSTRAP

This brief section on the DEA-estimation of the efficiency of production[15] starts with some basic definitions. A production set

(1) $$\Psi = \left\{ (x, y) \in \mathbb{R}_{+}^{p+q} \,\middle|\, x \text{ can produce } y \right\},$$

describes which amounts of some p inputs x can produce some q outputs y. An input requirement set $X(y)$ is defined as:

(2) $$X(y) = \left\{ x \in \mathbb{R}_{+}^{p} \,\middle|\, (x, y) \in \Psi \right\}$$

The standard assumptions w. r. t. these sets maintained here are a) the convexity of $X(y)$ for all y, b) that nonzero production of y requires some nonzero inputs x, and c) strong disposability of x and y. The efficient boundary of the input requirement set, $\partial X(y)$, is defined as:

(3) $$\partial X(y) = \left\{ x \,\middle|\, x \in X(y), \theta x \notin X(y) \;\forall\; 0 < \theta < 1 \right\}$$

and $\theta_k = \min\left\{ \theta \,\middle|\, \theta x_k \in X(y_k) \right\}$ is the input-oriented efficiency measure for a given combination of inputs and outputs, (x_k, y_k).

The sets Ψ and $X(y)$ as well as the efficient boundary $\partial X(y)$ are not observed but for

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

any given sample of observations $\mathcal{S} = \{(x_i, y_i) | i = 1,...,n\}$, the sample equivalents of (2), $\hat{X}(y)$, and (3), $\partial\hat{X}(y)$, as well as $\hat{\theta}$ can be derived. Specifically, $\hat{\theta}_k$ is obtained by solving[16]

(4) $$\hat{\theta}_k = \min\left\{\theta \middle| y_k \le \sum_{i=1}^{n} \lambda_i y_i ; \theta x_k \ge \sum_{i=1}^{n} \lambda_i x_i ; \theta > 0; \sum_{i=1}^{n} \lambda_i = 1; \gamma_i \ge 0, i = 1,...,n\right\}.$$

Non-zero weights λ_i are assigned to efficient producers on the frontier. These producers jointly constitute the reference technology. The condition $\Sigma_{i=1}^{n}\lambda_i = 1$ maintained in (4) leads to an evaluation based on a technology with variable returns to scale, i.e. the hospital that is evaluated and the reference technology operate on the same scale. Non-discretionary inputs can be included in the model by changing the condition $\theta x_k \ge \sum \lambda_i x_i$ in (4) to $x_k \ge \sum \lambda_i x_i$.

Efficiency estimates based on DEA-type methods are biased upwards. Since the observed frontier $\partial\hat{X}(y)$ can logically only be as good as the theoretical frontier $\partial X(y)$ but no better, the benchmark based on observations will in all likelihood be weaker than $\partial X(y)$; hence, the upward bias of the efficiency scores $\hat{\theta}$.

As can be seen from formula (4), the efficiency measure is calculated as the maximum proportional reduction of inputs for observation *k*, given that the benchmark units (the terms containing the λ_i) produce at least as much output with no more input than $\hat{\theta}_k x_k$.

To gain some intuition on how bias arises one may picture a situation where, in addition to the observations in this specific sample, other hospitals exist, which are not contained in the sample and at the same time are efficient when compared to the DMUs (decision making units: observations in a DEA) in the sample. Thus, the observed efficiency of inefficient producers calculated on the basis of what is observed is an upward biased estimate of their true efficiency.

Theoretical results on the sampling properties of the efficiency estimates are available only for the one-input-one-output case. Assuming a monotone, concave production function with a frontier function $g(\ )$ that is twice continuously differentiable at x_0 Simar and Wilson (2000, see section 3 and the results obtained by Gijbels et al., 1999, cited therein) derive the following expression[17] for the asymptotic bias

(5) $\quad \text{asymp. bias of } \hat{g}(x_0) = -n^{-2/3}\left(-g''(x_0)/2 \big/ f\left(x_0, g(x_0)\right)^2\right)^{1/3} c_1, \text{ where } \ c_1 = \text{constant}.$

This bias[18] depends on sample size n as well as on "the curvature of the frontier and the magnitude of the density at the frontier" (Simar and Wilson, 2000, p. 59). It should be intuitively clear that that the bias decreases in sample size and density and increases in curvature. That is, in large samples with a high density of observations around a frontier with a mild curvature, one should expect a relatively small bias; when the sample is small, the frontier exhibits kinks (changes in curvature) and the density of observations around the frontier is low, a relatively large bias is to be expected.

For the case of one input and output, it is possible to derive a bias corrected estimator on the basis of (5). The effect of the number of observations on bias will be even stronger for the case of multiple inputs and outputs. For this case, however, no expressions equivalent to (5) can be derived and in order to obtain bias corrected estimates for the multiple-input-multiple-output case, the bootstrap method must be applied. The estimates $\hat{\theta}_k$ and the bootstrap estimates $\hat{\theta}_k^*$ are related in the following way: $(\hat{\theta}_k - \theta_k)\big|\mathcal{S} \overset{approx.}{\sim} (\hat{\theta}_k^* - \hat{\theta}_k)\big|\mathcal{S}^*$. Thus, the bias of the DEA estimator in the general setting, $\text{bias}_{\mathcal{S},k} = E_{\mathcal{S}}(\hat{\theta}_k) - \theta_k$, can be estimated by its bootstrap counterpart $\widehat{\text{bias}}_{S^*,k} = E_{S^*}\left(\hat{\theta}_k\right) - \hat{\theta}$ and hence bias corrected estimates $\tilde{\theta}_k$ can be obtained with the correction $\tilde{\theta}_k = \hat{\theta}_k$ - $\widehat{\text{bias}}_{S^*,k} = 2\hat{\theta}_k - \bar{\theta}_k^*$, with $\bar{\theta}_k^* = R^{-1}\sum_R \bar{\theta}_k^*$. Of course, for a bias

For Peer Review

correction to be an improvement, the bias corrected estimator should not have a mean square error (MSE) larger than the ordinary estimates. For this to be the case, the condition $\mathrm{bias}_k^2/3 > \mathrm{var}(\theta_k^*)$ must hold (see Simar and Wilson, 2000).

IV Results

The results presented are representative for two important subsections of German hospitals. No representative results were obtained in previous studies (see the introductory section). Although some hospitals can potentially reduce input, i.e. lower per diem by almost 50%, the efficiency deficits on average range between 10% and 25% according to the bias corrected estimates.

The average score calculated with the bootstrap method (the standard DEA results in parentheses) for type I institutions is 0.75 (0.87) and 0.89 (0.94) for type II hospitals. The mean for all hospitals is 0.79 (0.89). Both differences of the two means were significant.[19] In all cases, the MSE test (see section III) was passed, which indicates the homogeneity of the samples.

The scores of the ordinary model are between the ones found by Steinmann et al. (2004) and Kuntz and Scholtes (2004). This is remarkable, as the hospitals in the dataset used for the present study are much more homogeneous than the ones used for the two German DEA studies just mentioned. This indicates that the efficiency of German hospitals was lower than previously thought because much of the inefficiency found in the two other studies could be attributed to the fact that the datasets comprised very different hospitals. The bias corrected figures cannot be compared to the results of the two other studies but they again indicate that the degree of inefficiency is much higher than previously thought.

None of the facilities in this study had a score below 50%, which would have been surprising given the homogeneity of the data. However, about half of the type I

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

hospitals are below 75 %. Only three of the type II facilities (or 6%) have a score this low. Since differences in case mix and hospital structure are controlled for it can be ruled out that the low efficiency of these units is due to an atypical case mix.

Table 3: Efficiency by group and ownership

The theoretical evidence on the relevance of the type of ownership for the efficiency of hospitals is mixed (see Burgess and Wilson, 1996; Mobley and Magnussen, 1998). German for-profit hospitals are thought to be significantly more efficient than other hospitals but there are too few of them in the data to allow any firm conclusions. Table 3 shows that efficiency does not differ significantly with the type of ownership for the hospitals in this sample.

Table 4 compares the characteristics of efficient and inefficient units. The average parameter values for the upper (efficient) quintile, the lower (inefficient) quintile and the three remaining quintiles are tabulated. Efficient hospitals in the upper quintile have lower per diem rates but LOS is not shorter in efficient hospitals compared to inefficient ones. Both the per diem rate and the mean cost per case (the latter variable is used merely for verification) differ significantly between the upper and lower quintile, mean LOS, however, is about the same.

For Peer Review

Table 4: Distribution of characteristics by efficiency score

Efficient regional-care hospitals, especially their surgery departments, have slightly longer LOS compared to their inefficient counterparts. The difference is, however, not significant. Nevertheless, some efficient hospitals have a higher average LOS in some

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

departments than inefficient ones highlighting the fact that isolated benchmarks are inadequate when inefficient institutions are to be identified. In this case, the criterion “duration of stay (in the surgery department)” would lead to efficient institutions being placed at the end of a duration-based ranking (one such study is Swart et al., 1997, see the introduction). Therefore, only a simultaneous benchmarking of all relevant variables (for instance by DEA) can lead to a useful assessment of efficiency.

Type I hospitals have a lower (standard) average score than type II hospitals. Given that the efficiency in both samples is assessed on the basis of the same parameters and since the expected upward bias is stronger the fewer observations are available it was to be expected that the dataset with more observations has a lower average efficiency.

However, the difference between the standard and the bootstrapped DEA results is much larger for type I hospitals, i.e. (relative) bias was actually stronger in the larger sample. For this to be the case, the observations in the larger sample must be more heterogeneous than the ones in the smaller sample. This may be interpreted in the following way: given the same case mix, type I hospitals have a less homogeneous structure of outputs and inputs, i.e. cases, LOS and cost than type II-hospitals. Whatever the effects that lead to this relatively higher degree of heterogeneity -these may be a lower qualification of the staff on which there is no information in the data, a lower learning rate due to the lower number of cases treated, etc.- it must be stronger than effects that may favour smaller institutions like the possibility of cream skimming (if they cream-skim at all) by referring more problematic cases to more advanced facilities.

V Conclusion

The conclusions derived in the present study are still largely valid today because the German hospital system has largely resisted attempts by policy makers to induce more efficiency. The main obstacle was that hospitals were entitled by law to full reimbursement of their cost; this was abolished only recently. The main finding of the

For Peer Review

study is that significant productivity differences between nearly identical hospitals exist. These differences are less dramatic than some findings in other studies on German hospitals; on the other hand, the bias-corrected results imply a much larger inefficiency than the results obtained in other DEA studies with German data. This is remarkable because the results of the present study were derived with rather homogenous data; consequently one may have expected smaller differences in relative efficiency compared to results obtained in other studies based on more heterogeneous samples.

The comparison of the average efficiency of the two samples showed that type I hospitals were on average less efficient than type II facilities. Therefore, the largest effects for the improvement of efficiency in the German hospital system could be expected if efforts followed a “bottom up” approach, concentrating on type I facilities

For Peer Review

NOTES

[1] The average of this figure was 5.2 during the 1960s, 7.9 during the 1970s and 9.1 during the 1980s. Taking some additional health care related expenditures into account, one seventh of the German GDP was spent on health care (OECD, 1997, p. 71). Sickness funds paid for about half of that. The latest figure is for 10.8% for 2002 (WHO, 2004).

[2] Occasionally, figures from 1994 are given instead of presenting up-to-date information because the data for the subsequent analysis are from this year. Most up-to-date figures are available in OECD (2001).

[3] Puig-Junoy (1998, p. 257) observes: “…, our study suggests that the most inefficient producers of health are Sweden, Denmark, Iceland, Germany and Finland”.

[4] The details of their methodology are given in Kuntz and Scholtes (2000).

[5] There are many advantages of using DEA rather than parametric alternatives for the evaluation of health care providers (see Steinmann and Zweifel, 2003)

[6] One study on hospitals efficiency using an alternative bootstrap procedure is Löthgren and Tambour (1999). For a discussion of some problems with their bootstrap approach, see Simar and Wilson (2000).

[7] Beginning in 1996, departmental per diem rates were introduced but no data are available.

[8] For an extensive description of the data source, see Arnold and Paffrath (1995) p. 273ff. and Arnold and Paffrath (1996), p. 279 ff.

[9] More than 2000 hospitals are contained in the 1995 raw data set and some 1800 in the 1996 data. There were 1700 hospitals that could be matched from both data sets. Due to the fact that reporting data was not mandatory at the time, the quality of the data was insufficient for some hospitals to be included in the 1996 report. The data in both reports are based on a 1994 sample of German hospitals.

[10] In many hospitals, a certain number of beds is reserved for patients of specialists who are not employed by the hospital but perform usually minor surgical procedures on in the hospital's facilities. The fraction of this type of beds is below 10% for group 11, between 10% and 20% for group 12 and exceeds 20% for group 13.

[11] The analysis therefore covers less than 10% of all German hospitals. It could be extended to some other hospitals with any of the five departments for which the information on LOS and case mix is available but the data sets would be smaller than the ones used here.

[12] See Chilingerian (2004) for some general comments on specification issues of DEA models used for evaluations in health care and Hollingsworth and Smith (2003) on ratio indicators.

[13] If capacity use was at 100% one of the two indicators would be redundant. But average capacity use in hospitals is well below 100%, varies considerably and exceeds 100% in some cases.

Unfortunately, cases per year are rounded to the accuracy of 1000 cases in the data which creates some imprecision. Therefore an alternative measure calculated as number of beds times days per year times average capacity use in the respective federal state (see Krankenhaus Report, 1995, table 17-11) is used to check the robustness of the results. No significant differences were found.

[14] Data kindly provided by Bettina Gerste, WIdO.

[15] Introductions into DEA can be found in a number of recent textbooks, e.g., Thanassoulis (2004) This section reiterates the DEA bootstrap set-up developed in Simar and Wilson (1998, 2000).

[16] Alternatively, one could solve the primal version of the linear programme. This so called multiplier form is often presented when weight restrictions are used. For a health application, see Womer et al. (2003).

For Peer Review

[17] This follows Simar and Wilson (2000) in giving expressions for the output oriented case.

[18] $b_0 = f\left(x_0, g\left(x_0\right)\right), \quad b_2 = -g''\left(x_0\right)/2, \quad c_1 = \text{constant}$ and $f\left(.\right)$: densitiy funtion

[19] For the standard DEA results, this is based on simple *t*-tests. For an evaluation of various tests regarding the significance of differences of efficiency scores, see Kittelsen (1999).

References

Arnold, M. and Paffrath D. (Eds.) (1995) Krankenhaus-Report ´95 Schwerpunkt: Neue Versorgungsstrukturen, Fischer, Stuttgart.

Arnold, M. and Paffrath D. (Eds.) (1996) Krankenhaus-Report ´96 Schwerpunkt: Managed Care, Fischer, Stuttgart.

Banker, R. D. and Morey R. C. (1986) Efficiency analysis for exogenously fixed inputs and outputs, *Operations Research*, **34**, 1613 - 27.

Burgess, J. F. and Wilson P. W. (1996) Hospital Ownership and Technical Inefficiency, *Management Science*, **42**, 110 – 23.

Chilingerian, J. A. and Sherman, H. D. (2004) Health Care Applications, in Handbook on Data Envelopment Analysis, (eds.) W. W. Cooper, L. M. Seiford, and J. Zhu, Kluwer, Boston, pp. 481 – 538.

Gijbels, I. E., Mammen, E., Park, U., and Simar, L. (1999) On Estimation of Monotone and Concave Production Functions, *Journal of the American Statistical Association*, **94**, 783 - 93

Gerste, B. (1996) Bildung von Krankenhausgruppen auf Fallmix-Basis, in *Krankenhaus-Report* ´96 (Eds.) M. Arnold and D. Paffrath, pp. 115–26.

Henke N., Paffrath D. and Wettke J. (1995). Benchmarking im Krankenhausmarkt, in *Krankenhaus-Report* ´95 (Eds.) M. Arnold and D. Paffrath, pp. 191 – 210.

Editorial Office, Dept of Economics, Warwick University, Coventry CV4 7AL, UK

For Peer Review

Hollingsworth, B., Dawson P. J. and Maniadakis, N. (1999) Efficiency Measurement of Health Care: A Review of Non-Parametric Methods and Applications, *Health Care Management Science* **2**, 151 – 72.

Hollingsworth, B. and Smith, P. (2003) Use of ratios in data envelopment analysis, *Applied Economics Letters* **10**, 733 - 735

Kittelsen, S. A. C. (1999) Monte Carlo simulations of DEA efficiency measures and hypothesis tests, Frisch Centre, Memorandum 9/1999.

Kuntz, L. und Scholtes, S. (2000) Measuring the Robustness of Empirical Efficiency Evaluations, *Management Science* **46**, 807 – 23.

Kuntz, L. und Scholtes, S. (2004) Effizienzanalyse zur Unterstützung der Krankenhausplanung, Arbeitsberichte zum Management in Gesundheitswesen Nr. 5.

Löthgren, M. and Tambour, M. (1999) Testing scale efficiency in DEA models: a bootstrapping approach, *Applied Economics* **31**, 1231 – 7.

Mobley, L. R. and Magnussen, J. (1998) An international comparison of hospital efficiency: does institutional environment matter? *Applied Economics* **30**, 1089–100.

OECD (1997) Economic Surveys: Germany, Paris.

OECD (2001) Health Data 2001, Paris.

Puig-Junoy, J. (1998) Measuring health production performance in the OECD, *Applied Economics Letters* **5**, 255 – 259.

Simar, L. and Wilson, P. (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Model, *Management Science*, **44**, 49 – 61.

Simar, L. and Wilson, P (2000) Statistical Inference in Nonparametric Frontier Models: the state of the art, *Journal of Productivity Analysis* **13**, 49 - 78.

Staat, M. (1999) Treating non-discretionary variables one way or the other: implications for efficiency scores and their interpretation, in Data Envelopment Analysis in the Service Sector (Ed.) G. Westermann, Gabler/DUV Wiesbaden.

For Peer Review

Steinmann, L., and Zweifel P. (2003) On the (in)efficiency of Swiss hospitals, *Applied Economics* **35**, 361 – 70.

Steinmann, L., Dittrich, G., Karmann, A. and Zweifel P. (2004) Measuring and Comparing the (In)Efficiency of German and Swiss Hospitals, *European Journal of Health Economics* **5**, 216 – 26.

Swart, E., Viethen G., Robra B.-P. and Leber, W.-D. (1996) Krankenhaus-Benchmarking anhand eines Verweildauerscores, *führen & wirtschaften im Krankenhaus* **13**, 226 – 29.

Thanassoulis, E. (2004). Introduction to the Theory and Application of Data Envelopment Analysis, Kluwer, Boston.

Womer, N. K., Shroff, H. F. E., Gulledge, T. R. and Haynes, K. E. (2003) Measuring efficiency with a linear economic model, *Applied Economics* **35,** 1459-1467.

WHO (2004) The World Health Report, Geneva.

For Peer Review

Table 1: Summary statistics

	Type I: local				Type II: regional			
Variable	mean	s. d.	min	max	mean	s. d.	min	max
per diem	393.6	51.7	260.6	574.8	389.8	50.6	310.4	581.7
avg. cost per case	4736	844	3500	8000	4606	1063	3000	10000
number of beds	162	44	70	265	196	73	50	441
cases treated p. a.	4907	1531	2000	8000	6192	2368	2000	14000
avg. LOS	11.39	1.86	7.5	18.4	11.26	1.80	7.9	20.5
avg. LOS internal med. dept.	12.39	1.95	7.3	19.1	11.89	2.01	7.8	16.5
avg. LOS surgery dept.	10.83	2.22	6.6	21.1	11.13	1.66	7.7	15.6
adverse case mix internal med.	.019	.14	0	1	.0577	.24	0	1
adverse case mix surgery	.046	.211	0	1	.212	.417	0	1
no. fields of specialisation	3.71	1.15	2	8	4.19	1.192	2	7

For Peer Review

Table 2: Characteristics of case mix clusters

Department Surgery	LOS in days	Intensive care days per 100 cases	Department Internal medicine	LOS in days	Intensive care days Per 100 cases	ICD-Cluster
Inner knee-joint	10.6	22.6	Low fraction of common ICDs.	13.8	41.7	1
No leading ICD.	11.2	30.0	No leading ICD	13.5	42.4	2
Hernia; cholelithiasis	11.3	28.1	Chronic cardiac condition	12.9	39.0	3
Arteriosclerosis; varicose veins.	11.4	33.1			39.9	4
Commotion; appendicitis	11.2	28.6				5
Rest	10.8	59.1	Rest	16.1		99

For Peer Review

Table 3: Efficiency by group and ownership

Structure group	Type I θ (σ) **obs.**	Type II θ (σ) **DMUs**	total θ(σ) **DMUs**	Owner-ship	Type I θ(σ) **DMUs**	Type II θ(σ) **DMUs**	total θ(σ) **DMUs**
	.74 (.10)	.87	.78	Non-	.73	.88	.77
11	**29**	(.07) **13**	(.11) **42**	profit	(.09) **53**	(.07) **18**	(.11) **71**
12	.75 (.07)	.89	.80	private	.74	.93	.78
	48	(.06) **23**	(.10) **71**		(.10) **04**	(.00) **01**	(.12) **05**
13	.76 (.09)	.89	.80	public	.77	.89	.82
	31	(.06) **16**	(.10) **47**		(.08) **51**	(.06) **33**	(.09) **84**
total	.75 (.09)	.88	.79	total	.75	.88	.79
	108	(.06) **52**	(.10) **160**		(.09) **108**	(.06) **52**	(.10) **160**

For Peer Review

Table 4: Distribution of characteristics by efficiency-score

μ (σ)	type I (basic) 0 – 20%	21 – 80%	81 – 100%	type II (standard) 0 – 20%	21 – 80%	81 – 100%
per diem rate	438.31	385.22	373.33	435.20	384.13	362.38
	(52.76)	(40.48)	(56.26)	(39.53)	(50.45)	(30.47)
avg. cost per case	5772.73	4585.94	4136.36	5800	4421.88	4000
	(1031.96)	(492.44)	(515.98)	(1619.33)	(623.59)	(577.35)
number of beds	157.77	165.47	158	218.9	186.47	203.9
	(35.36)	(44.51)	(53.47)	(56.34)	(66.29)	(105.08)
average LOS	12.77	11.14	10.76	12.74	10.89	10.98
	(2.24)	(1.62)	(1.51)	(2.91)	(1.21)	(1.30)
"Inner Medicine"	13.27	12.27	11.86	12.89	11.73	11.42
	(2.09)	(1.78)	(2.07)	(1.90)	(2.00)	(2.04)
"Surgery"	11.18	10.77	10.65	11.79	10.78	11.58
	(2.95)	(1.90)	(2.32)	(1.83)	(1.63)	(1.43)
Adverse mix of cases						
"Inner Medicine"	0	.02	.05	0	.06	.10
	(-)	(.13)	(.21)	(-)	(.25)	(.32)
"Surgery"	0	.02	.18	.30	.25 (.44)	0
	(-)	(.13)	(.39)	(.48)		(-)
Number of departments	2.95	3.78	4.27	3.90	4.19	4.5
	(.90)	(.97)	(1.49)	(1.20)	(1.18)	(1.27)
Observations	**22**	**64**	**22**	**10**	**32**	**10**

For Peer Review